Article:
Noyes, Eilidh and Jenkins, Rob orcid.org/0000-0003-4793-0435 (2019) Deliberate disguise in face identification. Journal of Experimental Psychology: Applied. ISSN 1076-898X

https://doi.org/10.1037/xap0000213

Deliberate disguise in face identification

Eilidh Noyes & Rob Jenkins

Department of Psychology, University of York, York, UK

Correspondence concerning this article should be addressed to:

Eilidh Noyes, Department of Psychology, School of Human and Health Sciences, University of Huddersfield, Queensgate, Huddersfield, HD1 3DH, UK

Email: e.noyes@hud.ac.uk

Phone: +44 1484 472770

Running Head: Deliberate disguise

Abstract

Facial image comparison is difficult for unfamiliar faces and easy for familiar faces. Those conclusions are robust, but they arise from situations in which the people being identified cooperate with the effort to identify them. In forensic and security settings, people are often motivated to subvert identification by manipulating their appearance, yet little is known about deliberate disguise and its effectiveness. We distinguish two forms of disguise—*evasion* (trying not to look like oneself) and *impersonation* (trying to look like another person). We present a new set of disguised face images (the FAÇADE image set), in which models altered their appearance to induce specific identification errors. In Experiment 1, unfamiliar observers were less accurate matching disguise items, and especially evasion items, than matching undisguised items. A similar pattern held in Experiment 2, in which participants were informed about the disguise manipulations. In Experiment 3, familiar observers saw through impersonation disguise, but accuracy was lower for evasion disguise. Quantifying the performance cost of disguise reveals distinct performance profiles for impersonation and evasion. Evasion disguise was especially effective, and reduced identification performance for familiar observers as well as for unfamiliar observers. We subsume these findings under a statistical framework of face learning.

Significance Statement

In security and forensic settings, individuals may be incentivized to alter their appearance to avoid identification. We show that (i) it is easier to avoid being recognized as oneself than to impersonate someone else, and (ii) disguises are less effective when viewers are familiar with the faces concerned.

Keywords: face recognition, face matching, identification, disguise, security

Introduction

Face images are frequently used to identify people. In criminal and security settings, this connection between facial appearance and personal identity can motivate people to alter their appearance, with a view to making identification more difficult. For example, one might transform one's usual appearance to avoid being recognised as oneself (*Evasion* disguise). Alternatively, one might imitate another individual, to pass oneself off as that person (*Impersonation* disguise). Some instances of disguise may reflect criminal intent—for example, a fugitive may wish to conceal his identity, or to mimic the photo in a stolen passport. However, identities can also be changed to protect people, as when ensuring the safety of undercover police or protected witnesses.

Applied face identification often relies on perceptual *face matching* (Bruce et al., 1999; Bruce, Henderson, Newman & Burton, 2001). A face-matching task involves comparing images of faces to establish whether they show the same person or different people. Twenty years of research has shown that accuracy on this task is surprisingly poor when the faces are unfamiliar to the viewers. Error rates are typically around 10–30% for images from different sources, even when overall image quality is good (e.g. Burton, White, & McNeill, 2010; Megreya & Burton, 2006; White, Kemp, Jenkins, Matheson & Burton, 2014). On the other hand, accuracy is very high when the faces are familiar (Clutterbuck & Johnston, 2002, 2004; Noyes & Jenkins, 2017; Jenkins, McLachlan, & Renaud, 2014), even when image quality is poor (Burton, Wilson, Cowan, & Bruce, 1999; Jenkins & Kerr, 2013). One way of understanding this familiarity contrast is through the statistics of perceptual experience. On this view, repeated exposure to a person's face builds up a representation of that particular face in the mind of the beholder (Jenkins & Burton, 2011). This mental representation encapsulates the range of variability seen for that person's appearance (Jenkins, White, Van Montfort, & Burton, 2011; Burton, Kramer, Ritchie, & Jenkins, 2016). A new image will be recognised as a specific person if it falls within the stored range for that particular person (Burton, Jenkins, & Schweinberger, 2011). The situation for an unfamiliar face is very different (Bruce, 1982; Burton & Jenkins, 2011). With no prior exposure to the unfamiliar face, the viewer has no representation of its variability. That makes it difficult to interpret differences between compared images. The differences might reflect variability within a single face, or they might reflect variability between faces. Without a larger sample of images (and the attendant face learning), it is hard to tell.

These principles are solid, and have clear implications for face identification in applied settings. However, they are derived almost entirely from studies in which the people being identified are cooperating with the effort to identify them. In virtually all of these studies, photographic models whose images were presented (Burton et al., 2010; White, Phillips, Hahn, Hill & O'Toole., 2015), or volunteers whose faces were viewed live (Kemp, Towell, & Pike, 1997; White et al. 2014), made no deliberate attempt to alter their facial appearance. Yet deliberate changes of appearance are attempted in some real world settings. Thus, projections from experimental evidence (no disguise) may underestimate the problem in those situations (disguise). Our purpose here is to use the image statistics framework outlined above to investigate deliberate disguise, and to calibrate expectations of face identification accuracy in the real world. We proceed with a brief summary of disguise research to date, and set out three opportunities to advance current understanding.

Among the first researchers to examine effects of disguise on face recognition performance were Patterson and Baddeley (1977). They compared recognition memory for learned face photos that were later presented in identical or disguised form. Recognition accuracy was higher for identical images than for disguised images. This finding may not be especially surprising, given that memory for specific face images tends to be good (Bruce, 1982). The basic result that disguise impairs recognition accuracy has since been replicated in several other memory experiments (Righi, Peissig, & Tarr, 2012; Terry, 1994; Terry, 1993).

One limitation of previous studies is that the techniques used to create disguise stimuli have often been simplistic. The most common approach has been to photograph models with and without a wig, fake beard, or glasses. Thus, the disguise amounts to adding standard paraphernalia that occludes parts of the underlying face (Patterson & Baddeley, 1977; Terry, 1993; Terry, 1994; Meissner, Susa, & Ross, 2013). In other studies, hairstyle changes, wigs, or glasses have been effected using computer graphics (Righi et al., 2012). Interestingly, Patterson and Baddeley (1977) took a rather different approach to disguise in their first experiment, using images of actors in different character roles as the learning and test images. Their approach has much to recommend it, as it leaves the means of disguise open-ended, rather than prescribing a particular means (such as donning glasses). It also allows different individuals to disguise themselves in different ways, rather than imposing the same generic change in appearance across all models. Despite these virtues, the Patterson and Baddeley (1977) approach to disguise did not catch on.

Second, the majority of research on disguise has overlooked the transformative effect of familiarity on face identification. Familiar face recognition is robust even in very challenging image conditions that reduce the accuracy of unfamiliar viewers (Burton et al., 1999; Clutterbuck & Johnston, 2002). This finding raises the question of whether familiar viewers see through disguise more easily than unfamiliar viewers. To our knowledge, Dhamecha, Singh, Vatsa, and Kumar (2014) are the only researchers to have compared familiar and unfamiliar viewers in the context of disguise. The main focus of their study was a comparison of human and computer performance for disguised faces. Viewers were presented with arrays of eight image pairs, most of which were disguised in some way (e.g. a face mask that obscured the mouth; sunglasses that hid the eyes). For each pair, the task was to decide whether the images showed the same person. Viewers who were familiar with the faces outperformed viewers who were not. However, the study could not directly compare performance for disguised and undisguised images, because both types of image were intermixed in the same displays.

Finally, previous disguise research has focused exclusively on *evasion* (trying not to look like oneself), and has overlooked *impersonation* (trying to look like a specific other person). This omission leaves a substantial disconnect between laboratory experiments and the real-world problem of identity fraud. In practice, identity fraud often involves use of another person's Photo-ID (Kemp et al., 1997; White et al., 2014), yet this important aspect of disguise has been entirely neglected.

The purpose of the current study was to address these three limitations—reliance on simplistic disguise manipulations, lack of data on familiarity effects, and absence of data on impersonation disguise. To this end, we first created a new image set —the FAÇADE image set—that documents the changes people made to their own appearance when motivated to induce identification errors. We then used FAÇADE images to construct a series of face matching experiments in which we compared identification accuracy for evasion and impersonation disguises, relative to undisguised images of the same people. Finally, we assessed effects of familiarity by testing viewers who were either unfamiliar (Experiments 1 & 2) or familiar (Experiment 3) with the faces in the FAÇADE image set.

The FAÇADE Image Set

To allow experimental manipulations of disguise, we constructed the FAÇADE image set, which is available from the authors as a research resource. The image set comprises photographs of volunteer models, each of whom was photographed in *Evasion*, *Impersonation* and *No Disguise* conditions. Rather than standardising changes in appearance, we sought to capture whatever changes the models deemed fit for purpose. For this reason, we specified disguise goals for the models (described below), but did not specify the means by which they were to achieve these goals. Our only stipulations were that (i) the end result should resemble a naturalistic face rather than someone in fancy dress, and (ii) that paraphernalia that would have to be removed for a passport security check (e.g. hats, scarves, masks) were not allowed.

Twenty-six early-career researchers (e.g. postgraduate students; postdoctoral researchers) and associates at the Department of Psychology, University of York volunteered as models (13 female, 13 male; age range 23–40). To create the different disguise conditions, we set disguise goals for these models based on reference photographs. For the *Evasion* condition, we gave the models their own work profile photograph, and asked them to make themselves look *unlike* that reference photo for a subsequent photo shoot. For the *Impersonation* condition, we gave them a reference photo of someone else (that other person's work profile photograph), and asked them to make themselves look *like* that person. We anticipated that the effectiveness of *Impersonation* disguises might depend on the initial similarity between the model's face and the target's face. Anecdotal accounts of identity theft suggest that fraudsters will pay a premium for photo-ID documents of people who look like them. To assess similarity effects, we created two impersonation conditions for each identity. *Impersonation Similar* paired the model with a similar-looking target (rated to be the most similar out of 33 potential match faces by a group of 3 viewers). *Impersonation Random* paired the model with a randomly chosen target of the same sex. To incentivise models to engage fully with the task, we introduced a cash reward, with the best disguise in each condition (based on participants' error rates in Experiment 1) receiving a £50 prize.

Disguise photo shoots were arranged separately for each model over a six-week period, allowing time for the models to plan and prepare their disguises, and for the experimenters to acquire any props and make-up requested. We also took a new *No Disguise* photograph of each model (in addition to the workplace photograph), to allow construction of undisguised same-identity pairs. Reference photos were taken between 6 months and 4 years before the experimental photo sessions (for comparison, passport photographs are often valid for up to 10 years). Figure 1 illustrates some of the possibilities.

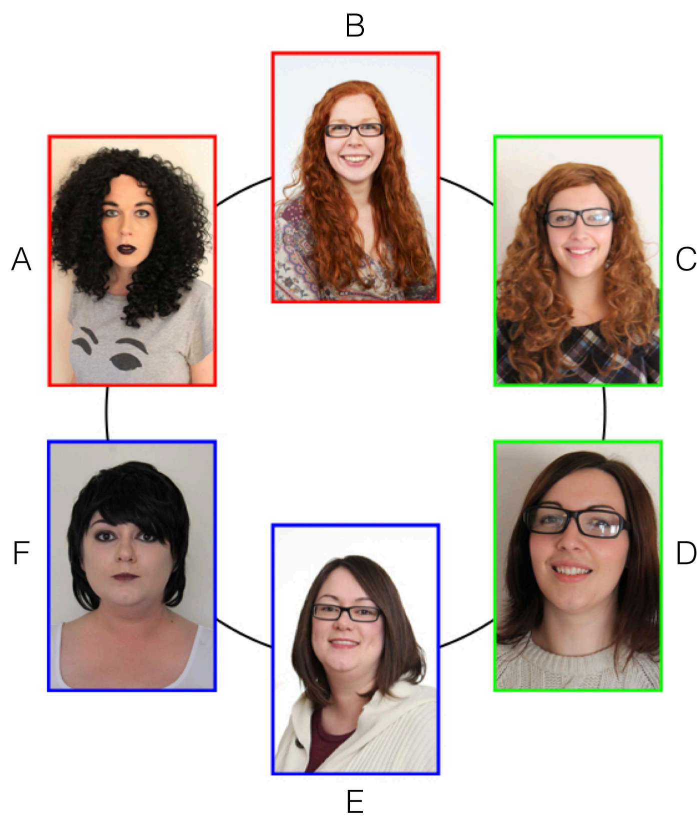

Figure 1. *Example images from the FAÇADE image set and the relations between them. Photos with the same colour frame show the same person. Photos with different colour frames show different people. (A) Model TF evasion disguise. (B) Model TF reference photograph. (C) Model HB impersonating TF. (D) Model HB impersonating KR. (E) Model KR reference photograph. (F) Model KR evasion disguise. Images A and F are unrelated, but share some features incidentally.*

Ethics statement

This study was approved by the Department of Psychology Ethics Committee at the University of York. All photographic models provided written informed consent for their images to contribute to the FAÇADE image set and to be used in psychology experiments. All participants in the experiments provided written informed consent.

Experiment 1

The purpose of Experiment 1 was to assess the impact of evasion and impersonation disguises on unfamiliar face matching performance. Unlike previous studies, we examined disguises that were freely chosen by the wearer to meet person-specific goals (to avoid one's own appearance, or to approach that of a particular target person). We expected that accuracy would be poorer overall for disguised faces than for undisguised faces. Of greater interest was whether viewers would be equally fooled by evasion disguise and impersonation disguise. Previous studies of unfamiliar face matching have shown that viewers find it difficult to integrate photos of the same face, even for natural samples of images (Jenkins et al., 2011; Andrews, Jenkins, Cursiter, & Burton, 2015). Presumably, deliberate efforts to disrupt that integration will make it harder still. For impersonation disguise, we also asked whether a person's similarity to the target matters. We expected that it might be easier for models to pass themselves off as targets whom they resembled anyway, compared with targets with a very different appearance. To address all of these questions, we presented photos of the same models in both *Disguise* and *No Disguise* pairs, so that we could isolate effects of different disguise types on matching accuracy.

The *No Disguise* condition was similar to a standard matching task for facial identity (Burton et al., 2010; White et al., 2015). Viewers were presented with pairs of face images and were asked to decide whether the two images showed the same person or two different people. In this case, we extended the standard design by including two types of different person trial. In *Different Similar* pairings, the foil was the most similar looking model in the FACADE image set. In *Different Random* pairings, the foil was drawn at random from models of the same sex. The *Disguise* condition repeated these identity pairings, but with models deliberately changing their appearance to meet the prescribed disguise goal. We expected that accuracy would be lower overall in the *Disguise* condition than the *No Disguise* condition. However, our main interest was the size of the performance decrement in each disguise category.

Method

Participants

Thirty undergraduate students from the University of York (mean age = 23 years; 22 female, 8 male) volunteered as participants in return for payment or course credit. All of the volunteers were naïve to the purpose of the experiment, and none of them was familiar with any of the models in the FAÇADE image set (confirmed by post-test familiarity check).

Design and Stimuli

Experimental stimuli were photo pairs constructed from the FAÇADE images (see Figure 2). The matching task used a 2 x 3 within-subjects design, with the factors *Disguise Condition* (*Disguise* and *No Disguise*) and *Pair Type* (*Same*, *Different Similar,* and *Different Random*). Each of the 26 models appeared in each of the six conditions, resulting in 156 trials in total. The dependent variable was percentage accuracy in the face-matching task for each of these conditions.

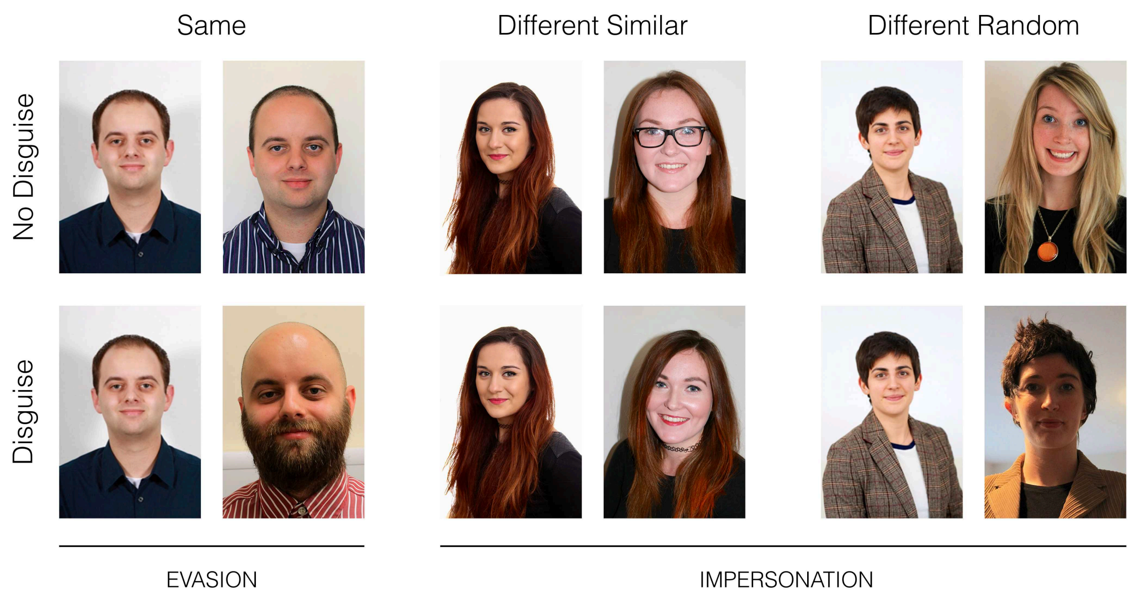

Figure 2. *Example image pairs for each condition of the face matching task in Experiment 1. The top row shows No Disguise images of the Same person (left), a Different person of Similar appearance (centre), and a Different person selected at Random (right). The bottom row shows the Disguise conditions, with the disguised image shown on the right of each pair. The Disguise could involve Evasion (in Same person pairs), or Impersonation (in Different person pairs). In Evasion disguise (left column) the disguised model (right image) tries to escape their normal appearance (left image). In Impersonation disguise (centre and right columns) a different model (right image) tries to mimic the normal appearance of the target model (left image).*

Procedure

In this experiment, participants were simply told they would be completing a face-matching task. Importantly, the possibility of disguise was not mentioned. Participants viewed the image pairs one at a time on a computer screen (viewing distance ~50cm). For each pair, the participants' task was to decide whether the two images showed the same person or two different people. Stimuli remained on screen until keypress response, and participants advanced through the task at their own pace without time pressure. Each participant viewed all 156 image pairs in a random order. Following the experiment, participants completed a familiarity check. Participants were presented with an image of each person in the stimulus set, and were asked to indicate whether or not they were familiar with each one. Items were categorised as familiar if the participant could name the person or provide individuating semantic information (e.g. person from lab X who studies Y). The entire task took approximately 20 minutes to complete.

Results

None of the participants was familiar with any of the faces in the experiment. Figure 3A shows the mean percentage accuracy rates across participants, separately for each of the six experimental conditions.

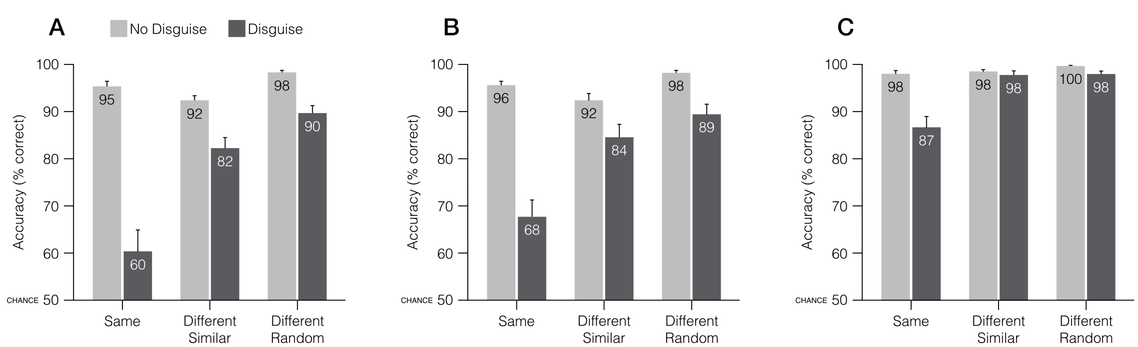

Figure 3. *Face matching accuracy (percent correct) for (A) unfamiliar viewers (Experiment 1), (B) informed unfamiliar viewers (Experiment 2), and (C) familiar viewers (Experiment 3). Each face appeared in Disguise (dark bars) and No Disguise (light bars), in Same, Different Similar, and Different Random identity pairings (x-axis; see main text for details). Error bars show the standard error of the mean.*

Note that a successful evasion disguise should increase the number of incorrect 'different person' responses in the *Same* condition, whereas a successful impersonation disguise should increase the number of incorrect 'same person' responses in the *Different Similar* and *Different Random* conditions.

As can be seen from the figure, accuracy was consistently lower in the *Disguise* condition than in the *No Disguise* condition. However, it was much lower for *Same* identity trials, in which evasion disguise incurred a 35% performance cost, compared to a 9% cost for impersonation. To analyse these differences, we submitted participants' mean accuracy scores to a 2 x 3 within-subjects ANOVA with the factors of *Disguise Condition (Disguise, No Disguise)* and *Pair Type (Same, Different Similar, Different Random)*. This analysis revealed a significant main effect of *Disguise Condition*, with lower accuracy overall for *Disguise* trials (M = 77%, SD = 21.27) compared with *No Disguise* trials (M = 95% accuracy, SD = 5.97) [F(1, 29) = 75.88, p<.001, η_p^2 = .72]. We also saw a significant main effect of *Pair Type* [F(2,58) = 22.87, p<.001, η_p^2 = .44], and a significant interaction between these two factors [F(2,58) = 37.95, p<.001, η_p^2 = .57].

Simple main effects confirmed that the effect of *Disguise Condition* was significant at each level of *Pair Type* (*Same* identity [F(1,87) = 151.5, p<.001, η_p^2 = .64], *Different Similar* [F(1,87) = 12.78, p<.001, η_p^2 = .13] and *Different Random* [F(1,87)= 9.19, p<.005, η_p^2 = .10]). The effect of *Pair Type* was significant in the *Disguise* condition [F(2,116) = 53.79, p<.001, η_p^2 = .48], but not in the *No Disguise* condition [F(2,116) = 2.03, p>.05, η_p^2 =.03].

Discussion

As expected, deliberate disguise impaired face-matching performance in this task. More interesting is that disguise *type* also made a difference. Evasion disguise (trying not to look like oneself) was much more effective than impersonation disguise (trying to look like a target individual). The performance decrement for impersonation was comparable for similar and randomly chosen disguise targets. One potentially important aspect of the procedure in Experiment 1 was that participants were not informed about the disguise manipulation until debrief. This raises the question of whether participants might be more adept at 'seeing through' disguises if they were fully briefed in advance. We addressed this possibility in the next experiment.

Experiment 2

In the previous experiment, deliberate disguise impacted accuracy in an identity matching task for unfamiliar faces. It is possible that this task was particularly difficult because participants were not expecting to see people in disguise. Expectations have been shown to influence accuracy in visual search and a range of other cognitive tasks (Reed, Ryan, McEntee, Evanoff & Brennan, 2011; Nocum, Brennan, Huang & Reed, 2013; Siipola, 1935; Wolfe, Horowitz, & Kenner, 2005). Based on such findings, we expected that if participants were informed about the disguise manipulations, the disguises might be less effective, leading to reduced error rates. To test this possibility, we recruited a new group of participants to complete the face-matching task presented in Experiment 1. The only difference was that we now explained to participants (i) the purpose of the study, (ii) the construction of the FAÇADE image set, and (iii) the design of the experiment, before presenting the task. As well as allowing direct comparison between informed and uninformed unfamiliar viewers, these conditions approximate everyday identification situations in which the use of disguise may be anticipated.

Methods

Participants

Thirty new undergraduate students at the University of York (mean age = 21 years; 19 female, 11 male) volunteered as participants in return for payment or course credit. None of these new participants had taken part in Experiment 1, and none of them was familiar with any of the models in the FAÇADE image set (confirmed by post-test familiarity check).

Design and Stimuli

The design and stimuli were the same as in Experiment 1.

Procedure

The procedure was the same as in Experiment 1, except that we now explained the design and stimuli to the participants during the task instruction. Thus, participants were aware from the start that half of the comparisons involved faces that were disguised to look unlike themselves or to look like another person.

Results

Summary data from Experiment 2 are presented in Figure 3B. As in Experiment 1, we submitted participants' accuracy data to a within-subjects ANOVA with the factors of *Disguise Condition* and *Pair Type*. This analysis again revealed a significant main effect for *Disguise Condition*, with participants performing more poorly with *Disguise* faces (M = 81%, SD = 18.68) than with *No Disguise* faces (M = 95%, SD = 6.62) [$F(1,29)= 44.25$, $p<.001$, $\eta_p^2 = .6$]. We also saw a significant main effect of *Pair Type* [$F(2,58)= 21.39$, $<.001$, $\eta_p^2 = .42$] and a significant interaction between *Disguise Condition* and *Pair Type* [$F(2,58) = 55.52$, $<.001$, $\eta_p^2 = .66$].

Simple main effects revealed a significant effect of *Disguise Condition* at every level of *Pair Type* (*Same* [$F(1,87) = 119.56$, $p<.001$, $\eta_p^2 = .58$], *Different Similar* [$F(1,87) = 9.45$, $p<.005$, $\eta_p^2 =.1$], and *Different Random* [$F(1,87) = 11.74$, $p<.001$, $\eta_p^2 = .12$]). On this occasion there was a significant effect of *Pair Type* in the *No Disguise* condition [$F(2,116) = 3.64$, $p<.01$, $\eta_p^2 = .06$] as well as the *Disguise* condition [$F(2,116) = 56.21$, $p<.001$, $\eta_p^2 = .49$], reflecting a small cost for similar disguise targets compared with randomly chosen disguise targets.

In sum, the pattern of results in Experiment 2 closely follows the pattern seen in Experiment 1. To formally test whether performance was any better for informed participants (Experiment 2) than for naïve participants (Experiment 1), we next carried out a between-experiments analysis of accuracy rates. There was no significant difference in overall accuracy between the two experiments [$F(1,58) = .83$, $p >.05$, $\eta_p^2 =.01$]. Indeed, Tukey HSD tests found no significant between-experiment differences for any of the six conditions ($p > .05$ for all comparisons).

Discussion

Accuracy in the unfamiliar face-matching task was again lower for disguised faces than for undisguised faces, even though participants were now informed about the disguise manipulation in advance. In addition, we replicated the finding that different types of disguise had different effects on performance, with evasion disguises attracting more errors than impersonation disguises. All of these results repeat the pattern established in

Experiment 1. In fact, pooling over Experiments 1 and 2 revealed no significant differences in accuracy across the two experiments. To summarise, briefing viewers on the use of disguises did not allow them to see through the disguises in this task. We next consider the role of familiarity in seeing through disguise.

Experiment 3

It is well established that perceptual experience with an individual's face improves identification of that face. In particular, familiar viewers can easily recognise individuals across changes in image that defeat unfamiliar viewers (Burton et al., 1999; Jenkins & Burton, 2011; Noyes & Jenkins, 2017). These changes may arise from variation in the face itself (e.g. expression, hairstyle; Young, McWeeny, Hay, Ellis, 1986; Young, Hay, McWeeny, Flude, & Ellis, 1985), variation in image properties (e.g. camera type, aspect ratio; Burton et al., 1999; Noyes & Jenkins, 2017), and variation in the conditions of image capture (e.g. lighting, viewpoint; Bruce, 1982; Hill & Bruce, 1996; O'Toole, Edelman, & Bülthoff, 1998). Given that familiar viewers are relatively impervious to incidental changes in appearance, it seems plausible that the same might be true for deliberate changes in appearance. If so, then viewers who were already familiar with the FAÇADE models before the experiment should perform relatively well on the matching task. Although this strikes us as a reasonable hypothesis, we do not regard it as a foregone conclusion. On the statistical view that we advance, a viewer's expertise with a particular face extends over the range of variability encountered. But it is also *bounded* by the range of variability encountered. That limitation explains why it can be difficult to recognise 'before they were famous' photos of well-known celebrities (unless the viewer knew them before they were famous; Russell, Duchaine, & Nakayama, 2009).

What is the status of disguised familiar faces in this scheme? The very purpose of disguise is to move the appearance of the face outside its normal range, and thus outside the range of viewers' perceptual experience. The question is whether this move makes it an unfamiliar face. If so, then familiar viewers should be no more accurate than unfamiliar viewers at matching disguised faces. On the other hand, if the disguise leaves enough information to engage the representation of a learned face, then familiar viewers should see through the disguise easily.

Methods

Participants

Thirty friends and colleagues of the models in the FAÇADE image set (mean age = 27 years; 16 female, 14 male) volunteered as participants in return for a small payment. All of these new participants were familiar with all of the FAÇADE models, as confirmed by a post-test familiarity check. None had taken part in the preceding experiments.

Design & Stimuli

The design and stimuli were the same as in Experiments 1 and 2, except for the participant group. In Experiment 3, participants were all personally familiar with the models in the FAÇADE image set.

Procedure

The procedure was the same as in Experiment 2. Participants were aware from the start that the faces they were comparing could be disguised to look unlike themselves or to look like another person.

Results

Summary data from Experiment 3 are presented in Figure 3C. As can be seen from the figure, accuracy approached ceiling in all conditions (~98%), with the exception of *Same Person* pairs in the *Disguise* condition (evasion disguise; 87%). As with the preceding experiments, we submitted accuracy data to a within-subjects ANOVA with factors of *Disguise Condition* and *Pair Type*. This analysis revealed a significant main effect of *Disguise Condition*, with lower accuracy overall for *Disguise* faces (94%) compared with *No Disguise* faces (99%) [$F(1,29) = 24.99$, $p < .001$, $\eta_p^2 = .46$]. There was also a significant main effect of *Pair Type* [$F(2,58) = 20.01$, $p < .001$, $\eta_p^2 = .41$], and a significant interaction between the two conditions [$F(2, 58) = 22.23$, $p < .001$, $\eta_p^2 = .43$], consistent with a disproportionate cost of disguise for *Same Person* pairs (evasion disguises).

Simple main effects confirmed that the effect of *Disguise* was significant for *Same Person* pairs [$F(1,87) = 68.44$, $p < .001$, $\eta p^2 = .44$], but not for *Different Similar* [$F(1,87) = 0.32$, p>.05, $\eta_p^2 < .01$] or *Different Random* pairs [F(1,87)= 1.73, p>.05, $\eta_p^2 = .02$]. The effect of *Pair Type* was significant for the *Disguise* condition [F(2,116) = 40.97, p<.001, $\eta_p^2 = .41$], but not for the *No Disguise* condition [F(2,116) = .73, p>.05, $\eta_p^2 = .01$].

Analysis of familiarity effects

To examine effects of *Familiarity* on this task, we directly compared performance across Experiment 2 (informed *Unfamiliar* participants) and Experiment 3 (informed *Familiar* participants). Accuracy data were submitted to a 2 x 2 x 3 mixed ANOVA, with the between-subject factor of *Familiarity*, and the within-subjects factors of *Disguise Condition* and *Pair Type*. This analysis revealed a main effect of *Familiarity* [F(1, 58) = 24.59, p<.001, $\eta_p^2 = .30$], with significantly higher accuracy overall when the faces were *Familiar* (Experiment 3; 96%) than when the faces were *Unfamiliar* (Experiment 2; 88%). The analysis also revealed significant main effects of *Disguise Condition* [F(1,58) = 64.97, p<.001, $\eta_p^2 = .53$], and *Pair Type* [F(2, 116) = 38.11, p<.001, $\eta_p^2 = .40$]. These main effects were modified by significant two-way interactions between *Familiarity* and *Disguise Condition* [F(1,58) = 17.86, p<.001, $\eta_p^2 = .24$], *Familiarity* and *Pair Type* [F(2,116) = 3.94, p<.05, $\eta_p^2 = .06$] and *Disguise Condition* and *Pair Type* [F(2, 116) = 76.65, p<.001, $\eta_p^2 = .57$]. There was also a significant three-way interaction between *Familiarity*, *Disguise Condition,* and *Pair Type* [F(2, 116) = 7.98, p <.005, $\eta_p^2 = .12$]. To break down this three-way interaction, we next carried out separate 2 x 3 mixed ANOVAs for the *No Disguise* and *Disguise* conditions, pooling data over Experiment 2 (*Unfamiliar*) and Experiment 3 (*Familiar*).

Effects of Familiarity on Disguise trials

The key question here is whether viewers see through disguises more easily when they are familiar with the disguised faces, compared with viewers who are unfamiliar with the disguised faces. Analysis of *Disguise* trials revealed a significant main effect of *Familiarity* [F(1, 58) = 22.94, p<.001, $\eta_p^2 = .28$], with higher overall accuracy for *Familiar* participants (94%) than for *Unfamiliar* participants (81%). There was also a significant main effect of *Pair Type* [F(2, 116) = 59.68, p<.001, $\eta_p^2 = .51$], with lower accuracy for evasion disguise (*Same Person* condition, 77%) than for impersonation disguise (*Different Similar* condition, 91%; *Different Random* condition, 94%). These main effects were modified by a significant interaction between *Familiarity* and *Pair Type* [F(2, 116) = 5.29, p<.01, η_p^2 = .08].

Simple main effects revealed a significant effect of *Familiarity* for both evasion disguise (*Same Person* trials) [F(1,174) = 31.33, p<.001, η_p^2 = .15], and impersonation disguise (*Different Similar* [F(1,174) = 15.17, p<.001, η_p^2 = .08]; *Different Random* [F(1,174) = 6.23, p<.01, η_p^2= .03]). There was also a main effect of *Pair Type* for both *Unfamiliar* viewers [F(2,116) = 49.32, p<.001, η_p^2 = .46] and *Familiar* viewers [F(2,116) = 15.65, p<.001, η_p^2 = .21].

Effects of Familiarity on No Disguise trials

For completeness, we carried out a similar analysis of *No Disguise* trials. This analysis found a main effect of *Familiarity* [F(1, 58) = 15.81, p<.001, η_p^2 = .21], with *Familiar* participants performing at higher overall accuracy (98%) than *Unfamiliar* participants (95%). There was also a significant main effect of *Pair Type* [F(2, 116) = 8.96, p<.001, η_p^2 = .13] with lower accuracy for *Different Similar* pairs (95%) than for *Same Person* (97%) or *Different Random* pairs (99%), although the differences were numerically small. These main effects were modified by a significant interaction between *Familiarity* and *Pair Type* [F(2, 116) = 5.29, p<.01, η_p^2 = .08].

Simple main effects revealed a significant effect of *Familiarity* for *Different Similar* trials [F(1,174) = 23.28, p = <.001, η_p^2 = .12] but not for *Same Person* [F(1,174) = 3.65, p = >.05, η_p^2 = .02] or *Different Random* trials [F(1,174) = 1.45, p = >.05, η_p^2= .01]. There was also a significant effect of *Pair Type* for *Unfamiliar* viewers [F(2,116) = 12.3, p = <.001, η_p^2 = .17], but not for *Familiar* viewers [F(2,116) = 1.07, p = >.05, η_p^2 = .02].

Discussion

Accuracy in the disguised face-matching task was relatively high for viewers who knew the people being disguised. In fact, for familiar viewers, accuracy for impersonation disguises was as high as for undisguised faces. However, even familiar viewers were impaired by evasion disguises (11% cost), albeit to a lesser extent than informed unfamiliar viewers (28% cost). Although many studies have shown that familiar viewers are impervious to many incidental variations in facial appearance, deliberate disguise can move the appearance of a known face outside the recognisable range. We return to this point in our General Discussion.

General Discussion

We constructed the FAÇADE image set to facilitate experimental investigation of deliberate disguise. This new image set, comprising photographs of models in disguised and undisguised form, builds on earlier work in several important ways. Previous stimulus sets have been limited to evasion disguise, and have applied the same, standardised manipulations to each face (e.g. Righi, Peissig, & Tarr, 2012; Terry, 1993; Terry, 1994). The FAÇADE image set captures idiosyncratic disguises, devised by the models themselves to meet evasion and impersonation goals. Unlike other disguise image sets, it includes both male and female faces, rather than males only. It also excludes props and paraphernalia that would violate everyday face identification guidelines by occluding facial features. The FAÇADE image set is available from the authors as a resource for face perception research.

Rather than focusing on recognition memory performance, we presented FAÇADE images in a series of perceptual matching experiments, providing a direct comparison between participants who were familiar versus unfamiliar with the FAÇADE models. In Experiment 1, unfamiliar viewers performed relatively accurately when the faces were undisguised (cf. Burton, White, & McNeill, 2010), presumably because the number of possible identities was small. As expected, accuracy was much lower when the faces were disguised. Moreover, the accuracy cost was disproportionately large for evasion disguise, in which models sought not to be recognised as themselves. A similar pattern emerged in Experiment 2, in which unfamiliar viewers were briefed on the disguise manipulations before testing. Awareness of this manipulation did not reduce the rate of errors or alter their distribution across conditions. In Experiment 3, viewers who were familiar with the models' faces performed at ceiling when the faces were undisguised, consistent with previous studies of familiar face recognition (Bruce & Young, 1986; Burton et al., 1999; Jenkins et al., 2011; Noyes & Jenkins, 2017). Interestingly, familiar viewers also readily saw through the impersonation disguises. However, even familiar viewers could be fooled by evasion disguises. A few previous studies have probed the limits of familiar face matching by presenting very low quality images (e.g. Robertson, Noyes, Dowsett, Jenkins & Burton, 2016; Jenkins & Kerr, 2013). The present case is the first that we are aware of to approach those limits using normal quality photographs.

How might this pattern of findings be explained? Two points seem to us important. First, it is well established that internal facial features (eyes, nose and mouth) play a key role in familiar face recognition, whereas recognition of unfamiliar faces is often very dependent on external features such as hairstyle (Clutterbuck & Johnston, 2002; Ellis, Shepherd, & Davies, 1979; Young et al., 1985). The FAÇADE models were clearly very creative in their approaches to disguise (see Figure 1). But without the services of a professional makeup artist or a plastic surgeon, the scope for changing the appearance of internal features was more limited than the scope for changing hairstyle and hair colour. This limited ability to alter the appearance of internal features may explain failures of impersonation disguises in Experiment 3, although we can not rule out other factors with the current experimental design.

Our second point concerns the nature of face recognition itself. Recent computational and theoretical approaches show how some of the problems that image variability presents in everyday face recognition can be solved by creating a perceptual space that is reshaped around regions corresponding to familiar identities (Jenkins & Burton, 2011; Kramer, Young, & Burton, 2018; Kramer, Young, Day, & Burton, 2017). This conceptual framework naturally accommodates the observation that familiar viewers performed more accurately than unfamiliar viewers in this task. The proposal is that exposure to a person's face refines the viewer's mental representation of that face, so that the representation comes to incorporate the range of variability in appearance for that person (Bruce, 1994; Burton, Jenkins, Hancock, White, 2005; Burton, Kramer, Ritchie, Jenkins, 2016). For an unfamiliar viewer, even minor changes in appearance might constitute an effective disguise, as the acceptable range of appearance for that identity is not known. But for a familiar viewer, the same changes in appearance may fall within the acceptable range, as that range has been informed by prior perceptual experience with the face in question. We suggest that a better grasp of the within-person variability for each face supports more accurate identity judgements.

The same framework can explain why evasion disguises led to more errors than impersonation disguises: there are many more ways to leave one's own region of facial appearance than to enter someone else's. For example, changing your hair colour might make you look less like your usual self, supporting the goal of evasion. But it would only support the goal of impersonation if your new hair colour matches that of the target. The particular difficulty in seeing through evasion disguise echoes the more general problem of 'telling people together' in normal face recognition (Jenkins et al., 2011; Andrews, Jenkins, Cursiter, & Burton, 2015). In the context of incidental image variability, unfamiliar viewers often find telling people together harder than telling people apart, whereas familiar viewers have little trouble with either. The demands of telling people together and seeing through evasion disguise are closely related, in that both require mapping different images onto the same identity. Evasion disguise presents a particularly extreme case because divergences in appearance are deliberate rather than incidental, and are calculated to mislead. If a person changes their facial appearance outside of its normal range, mastery of that normal range might not be much help.

In discussing these findings, we have been confident in ascribing higher accuracy in Experiment 3 to higher familiarity of the participants with the target faces. A reviewer asked whether other differences between experiments might account for the better performance. For example, the participants in Experiment 3 were not only four years older on average, they had also spent more time in psychology education. Perhaps they had acquired expert knowledge about face identification that helped them to solve the task. Age seems unlikely to account for the observed differences, given that face processing ability is known to be stable across the age range that we tested. Germine, Duchaine, & Nakayama (2010) tested 44,000 10–70 year olds on the Cambridge Face Matching Test, and found an increase in accuracy of just 2% from early- to late-twenties. Susilo, Germine, & Duchaine (2013) reported similar findings. This brings us to possible effects of additional psychology education. There is a superficial response to this concern, which is that participants in Experiment 3 had not in fact acquired expert knowledge about face identification and disguise. Those topics were not part of the postgraduate taught programme, and none of the participants was conducting research in those areas. However, there is a deeper point too, which is that gaining expert knowledge about face identification does not seem to improve performance. In a recent demonstration of this, White et al. (2014) found that trained passport officers were no more accurate than undergraduate students at matching unfamiliar faces. Moreover, there was no correlation between time in post (0–20 years) and accuracy. Papesh (2018) reported similar findings based on other practitioner groups. Even training programmes that were specifically designed to enhance face identification performance appear to confer little or no benefit (e.g. Towler, White & Kemp, 2014, 2017). In short, despite the tremendous applied incentive to reduce identification errors, only one factor has been found to bring about meaningful improvements in performance, and that is familiarity with the faces concerned (see Johnston & Edmonds, 2009; Jenkins & Burton, 2011; Burton, Jenkins & Schweinberger, 2011 for reviews).

To achieve a strong manipulation of familiarity in this study, we used a between-groups design in which different participants were recruited as familiar viewers and unfamiliar viewers. Familiar viewers already knew the disguised individuals before the experiment, through day-to-day social exposure. Unfamiliar viewers had never seen the disguised individuals before. This approach has the advantages of capturing familiarity differences as they occur in real life, and allowing massive disparities in prior exposure (compared with, say, learning experiments in which participants are 'familiarized' with new faces as part of the study). The disadvantage of this between-groups design is that it can not definitively rule out other between-groups differences as contributing factors (as discussed above). We interpret elevated accuracy in Experiment 3 in the light of converging evidence accumulated over several decades.

Previous studies of unfamiliar face matching have consistently found high error rates, even when there is no intent to deceive. Here we found that error rates were much higher in the context of deliberate disguise. The implications for applied facial image comparison are clear. For situations in which people are motivated to change their appearance, estimates of identification performance that ignore that motivation probably underestimate identification errors. The reasonable assumption that such situations are rare offers little comfort, as rare occurrences are more frequently missed in vigilance tasks (Wolfe, Horowitz, & Kenner, 2005; Wolfe, Brunelli, Rubenstein, & Horowitz, 2013). We had expected that informing participants about the disguise manipulation might improve their performance (Experiment 2). As it turned out, there was no appreciable benefit to being informed. If performance on this task is to be improved, other possibilities will have to be explored. It is clear that familiar viewers can achieve higher levels performance (Experiment 3). That might be a practical solution in some cases, but it is difficult to see how it could scale. Training programmes designed to enhance facial image comparison have generally had disappointing results (e.g. Towler, White & Kemp 2014, 2017). It is possible that knowledge of the methods people use to disguise themselves might enhance performance on this task, but we are not especially optimistic about this possibility. Even in the current study, different individuals used different methods of disguise for different purposes. This makes it difficult to see how useful generalisations could be extracted. Crowd analysis—the method of pooling independent decisions from multiple viewers—has been shown to improve face matching performance in previous work (White Burton, Kemp, & Jenkins, 2013; Jeckeln, Hahn, Noyes, Cavazos, & O'Toole, 2018). Given the generality of crowd effects (Galton, 1907; Ariely et al., 2000; Surowiecki, 2004), there are good reasons to expect that crowd analysis could improve identification accuracy for disguised faces too. Finally, computer algorithms play an important role in applied face recognition (see Phillips & O'Toole, 2014; Phillips, 2017 for recent reviews). Given that computer algorithms can diverge from human information processing, it would be informative to compare accuracy across systems and look for analogous or complementary patterns.

In summary, the current experiments advance our understanding of facial disguise in several ways. The simple observation that disguise impairs face matching implies that, in some situations, performance estimates based on undisguised faces may be overoptimistic. Studies that rely on generic disguises mischaracterise the problem, because people in fact deploy idiosyncratic disguises to meet specific goals. Disguises were less effective for viewers who knew the faces, compared with viewers who did not know the faces. Evasion disguise affects face-matching accuracy more strongly than impersonation disguise—a finding that holds for familiar and unfamiliar faces alike. These distinctions will have to be acknowledged if deliberate disguise in applied settings is to be confronted.

Acknowledgements

We thank those members of the Department of Psychology, University of York who volunteered as photographic models in this research, and Andy Young for helpful discussions of the findings.

References

Andrews, S., Jenkins, R., Cursiter, H., & Burton, A. M. (2015). Telling faces together: Learning new faces through exposure to multiple instances. *The Quarterly Journal of Experimental Psychology*, *68*, 1–10. https://doi.org/10.1080/17470218.2014.1003949

Ariely, D., Tung Au, W., Bender, R. H., Budescu, D. V., Dietz, C. B., Gu, H., Wallsten, T.S, & Zauberman, G. (2000). The effects of averaging subjective probability estimates between and within judges. *Journal of Experimental Psychology: Applied*, *6*(2), 130–147.

Bruce, V. (1982). Changing faces: Visual and non-visual coding processes in face recognition. *British Journal of Psychology*, *73*, 105–116.

Bruce, V. (1994). Stability from variation: The case of face recognition the MD Vernon memorial lecture. *The Quarterly Journal of Experimental Psychology*, *47*(1), 5-28.

Bruce, V., Henderson, Z., Greenwood, K., Hancock, P. J., Burton, A. M., & Miller, P. (1999). Verification of face identities from images captured on video. *Journal of Experimental Psychology: Applied*, *5*(4), 339.

Bruce, V., Henderson, Z., Newman, C., & Burton, A. M. (2001). Matching identities of familiar and unfamiliar faces caught on CCTV images. *Journal of Experimental Psychology: Applied*, *7*(3), 207.

Bruce, V., & Young, A. (1986). Understanding face recognition. *British Journal of Psychology*, *77*, 305–327.

Burton, A. M., & Jenkins, R. (2011). Unfamiliar face perception. *The Oxford handbook of face perception*, 287-306.

Burton, A. M., Jenkins, R., Hancock, P. J., & White, D. (2005). Robust representations for face recognition: The power of averages. *Cognitive Psychology*, *51*(3), 256-284.

Burton, A. M., Jenkins, R., & Schweinberger, S. R. (2011). Mental representations of familiar faces. *British Journal of Psychology*, *102*(4), 943-958.

Burton, A. M., Wilson, S., Cowan, M., & Bruce, V. (1999). Face Recognition in Poor-Quality Video: Evidence From Security Surveillance. *Psychological Science*, *10*(3), 243–248. https://doi.org/10.1111/1467-9280.00144

Burton, A M., White, D., & McNeill, A. (2010). The Glasgow Face Matching Test. *Behavior Research Methods*, *42*, 286–91.

Burton, A. M., Kramer, R. S., Ritchie, K. L., & Jenkins, R. (2016). Identity from variation: Representations of faces derived from multiple instances. *Cognitive Science*, *40*(1), 202-223.

Clutterbuck, R., & Johnston, R. A. (2002). Exploring levels of face familiarity by using an indirect face-matching measure. *Perception*, *31*(8), 985–994. https://doi.org/10.1068/p3335

Clutterbuck, R. & Johnston, R.A. (2004). Matching as an index of face familiarity. *Visual Cognition, 11*(7), 857-869.

Dhamecha, T. I., Singh, R., Vatsa, M., & Kumar, A. (2014). Recognizing disguised faces: Human and machine evaluation. *PLoS ONE*, *9*(7). https://doi.org/10.1371/journal.pone.0099212

Ellis, H. D., Shepherd, J. W., & Davies, G. M. (1979). Identification of familiar and unfamiliar faces from internal and external features: some implications for theories of face recognition. *Perception*, *8*(4), 431–9. Retrieved from http://www.ncbi.nlm.nih.gov/pubmed/503774

Galton, F. (1907). Vox populi (The wisdom of crowds). *Nature*, *75*(7), 450-451.

Germine, L. T., Duchaine, B., & Nakayama, K. (2011). Where cognitive development and aging meet: Face learning ability peaks after age 30. *Cognition, 118*(2), 201-210.

Hill, H., & Bruce, V. (1996). Effects of lighting on the perception of facial surfaces. *Journal of Experimental Psychology. Human Perception and Performance*, *22*(4), 986. https://doi.org/10.1037/0096-1523.22.4.986

Jeckeln, G., Hahn, C. A., Noyes, E., Cavazos, J. G., & O'Toole, A. J. (2018). Wisdom of the social versus non-social crowd in face identification. *British Journal of Psychology*. https://doi.org/10.1111/bjop.12291

Jenkins, R., & Burton, A. M. (2011). Stable face representations. *Philosophical Transactions of the Royal Society of London. Series B, Biological Sciences*, *366*(1571), 1671–1683. https://doi.org/10.1098/rstb.2010.0379

Jenkins, R., & Kerr, C. (2013). Identifiable images of bystanders extracted from corneal reflections. *PloS one*, *8*(12), e83325.

Jenkins, R., McLachlan, J. L., & Renaud, K. (2014). Facelock: familiarity-based graphical authentication. *PeerJ*, *2*, e444.

Jenkins, R., White, D., Van Montfort, X., & Mike Burton, A. (2011). Variability in photos of the same face. *Cognition*, *121*, 313–23.

Johnston, R. A., & Edmonds, A. J. (2009). Familiar and unfamiliar face recognition: A review. *Memory*, *17*(5), 577-596.

Kemp, R., Towell, N., & Pike, G. (1997). When seeing should not be believing: Photographs, credit cards and fraud. *Applied Cognitive Psychology*, *11*(3), 211-222.

Kramer, R. S. S., Young, A. W., & Burton, A. M. (2018). Understanding face familiarity. *Cognition*, *172*, 46–58.

Kramer, R. S. S., Young, A. W., Day, M. G., & Burton, M. A. (2017). Robust social categorization emerges from learning the identities of very few faces. *Psychological Review*, (124), 115–129.

Megreya, A. M., & Burton, A. M. (2006). Unfamiliar faces are not faces: Evidence from a matching task. *Memory & Cognition*, *34* (4), 865-876.

Meissner, C. A., Susa, K. J., & Ross, A. B. (2013). Can I see your passport please? Perceptual discrimination of own-and other-race faces. *Visual Cognition*, *21*(9-10), 1287-1305.

Nocum, D. J., Brennan, P. C., Huang, R. T., & Reed, W. M. (2013). The effect of abnormality-prevalence expectation on naïve observer performance and visual search. *Radiography*, *19*(3), 196-199.

Noyes, E., & Jenkins, R. (2017). Camera-to-subject distance affects face configuration and perceived identity. *Cognition*, *165*, 97–104. https://doi.org/10.1016/j.cognition.2017.05.012

O'Toole, A. J., Edelman, S., & Bülthoff, H. H. (1998). Stimulus-specific effects in face recognition over changes in viewpoint. *Vision Research*, *38*, 2351–2363.

Papesh, M. H. (2018). Photo ID verification remains challenging despite years of practice. *Cognitive Research: Principles and Implications*, *3*(1), 19.

Patterson, K. E., & Baddeley, A. D. (1977). When face recognition fails. *Journal of Experimental Psychology. Human Learning and Memory*, *3*(4), 406–17. Retrieved from http://www.ncbi.nlm.nih.gov/pubmed/864389

Phillips, P. J. (2017). A Cross Benchmark Assessment of a Deep Convolutional Neural Network for Face Recognition. In *Automatic Face & Gesture Recognition (FG 2017), 2017 12th IEEE International Conference on* (pp. 705-710). IEEE.

Phillips, P. J., & O'Toole, A. J. (2014). Comparison of human and computer performance across face recognition experiments. *Image and Vision Computing*, *32*(1), 74-85.

Reed, W. M., Ryan, J. T., McEntee, M. F., Evanoff, M. G., & Brennan, P. C. (2011). The effect of abnormality-prevalence expectation on expert observer performance and visual search. *Radiology*, *258*(3), 938-943.

Righi, G., Peissig, J. J., & Tarr, M. J. (2012). Recognizing disguised faces. *Visual Cognition*, *20*(2), 143–169. https://doi.org/10.1080/13506285.2012.654624

Robertson, D. J., Noyes, E., Dowsett, A. J., Jenkins, R., & Burton, A. M. (2016). Face recognition by metropolitan police super-recognisers. *PloS one*, *11*(2), e0150036.

Russell, R., Duchaine, B., & Nakayama, K. (2009). Super-recognizers: People with extraordinary face recognition ability. *Psychonomic Bulletin & Review*, *16*, 252–257.

Siipola, E. M. (1935). A group study of some effects of preparatory set. *Psychological Monographs*, *46*(6), 27.

Surowiecki, J. (2004). The wisdom of crowds. 2004. *New York: Anchor.*

Susilo, T., Germine, L., & Duchaine, B. (2013). Face recognition ability matures late: Evidence from individual differences in young adults. *Journal of Experimental Psychology: Human Perception and Performance, 39*(5), 1212.

Terry, R. L. (1993). How wearing eyeglasses affects facial recognition. *Current Psychology*, *12*(2), 151–162. https://doi.org/10.1007/BF02686820

Terry, R. L. (1994). Effects of facial transformations on accuracy of recognition. *The Journal of Social Psychology*. https://doi.org/10.1080/00224545.1994.9712199

Towler, A., White, D., & Kemp, R. I. (2014). Evaluating training methods for facial image comparison: The face shape strategy does not work. *Perception*, *43*(2–3), 214–218. https://doi.org/10.1068/p7676

Towler, A., White, D., & Kemp, R. I. (2017). Evaluating the feature comparison strategy for forensic face identification. *Journal of Experimental Psychology: Applied*, *23*(1), 47.

White, D., Burton, A. M., Kemp, R. I., & Jenkins, R. (2013). Crowd Effects in Unfamiliar Face Matching, *27*, 769–777.

White, D., Kemp, R. I., Jenkins, R., Matheson, M., & Burton, A. M. (2014). Passport officers' errors in face matching. *PloS One*, *9*, e103510.

White, D., Phillips, P. J., Hahn, C. A., Hill, M., & O'Toole, A. J. (2015). Perceptual expertise in forensic facial image comparison. *Proc. R. Soc. B*, *282*(1814), 20151292.

Wolfe, J. M., Horowitz, T. S., & Kenner, N. M. (2005). Cognitive psychology: rare items often missed in visual searches. *Nature*, *435*(7041), 439.

Wolfe, J. M., Brunelli, D. N., Rubinstein, J., & Horowitz, T. S. (2013). Prevalence effects in newly trained airport checkpoint screeners: Trained observers miss rare targets, too. *Journal of vision*, *13*(3), 33-33.

Young, A. W., Hay, D. C., McWeeny, K. H., Flude, B. M., & Ellis, A. W. (1985). Matching familiar and unfamiliar faces on internal and external features. *Perception*, *14*(6), 737–46. Retrieved from http://www.ncbi.nlm.nih.gov/pubmed/3837875

Young, A. W., McWeeny, K. H., Hay, D. C., & Ellis, A. W. (1986). Matching familiar and unfamiliar faces on identity and expression. *Psychological Research*, *48*(2), 63-68.